Multivariate Normal & Wishart

Hoff Chapter 7

October 14, 2009
Twenty-two children are given a reading comprehension test before and after receiving a particular instruction method.

- $Y_{i,1}$: pre-instructional score for student $i$
- $Y_{i,2}$: post-instructional score for student $i$
- vector of observations for each student $\mathbf{Y}_i = (Y_{i,1}, Y_{i,2})'$

Questions:

- Does the instructional method lead to improvements in reading comprehension (on average)?
- If so, by how much?
- Can improvements be predicted based on the first test?
Parameters

We could model the data as bivariate normal \( Y_i \overset{iid}{\sim} N_2(\mu, \Sigma) \). Normal distributions are characterized by their

- mean vector \( \mu = (\mu_1, \mu_2)' \) where \( \mu_1 = \mathbb{E}[Y_{i,1}] \) and \( \mu_2 = \mathbb{E}[Y_{i,2}] \)
- variance-covariance matrix

\[
\Sigma = \begin{bmatrix}
\sigma_1^2 & \sigma_{12} \\
\sigma_{21} & \sigma_2^2 \\
\end{bmatrix}
\]

where \( \sigma_1^2 \) is the variance of \( Y_{i,1} \), \( \sigma_2^2 \) is the variance of \( Y_{i,2} \), and \( \sigma_{12} \) is the covariance between \( Y_{i,1} \) and \( Y_{i,2} \),

\[
\sigma_{21} = \sigma_{12} = \mathbb{E}[(Y_{i,1} - \mu_1)(Y_{i,2} - \mu_2)]
\]

- Correlation between pre and post test scores is \( \rho = \frac{\sigma_{12}}{\sigma_1 \sigma_2} \), measure of the strength of association.
Need to specify prior distributions, then use Bayes Theorem to obtain posterior distributions.

- Normal $N(\mu_0, S_0)$ is the conjugate prior for the mean (given the covariance matrix $\Sigma$).
- Easier to work with the inverse of $\Sigma$, $\Phi \equiv \Sigma^{-1}$
- Given the mean vector, the conjugate prior for $\Phi$ is the Wishart distribution, a generalization of the gamma distribution to higher dimensions.
- If $\Phi \sim \text{Wishart}(\nu, \Phi_0)$ then
  - $E[\Phi] = \nu \Phi_0$
  - $E[\Sigma] = \frac{1}{\nu - p - 1} \Phi_0^{-1}$
Prior for $\mu$:

$$\mu \sim N_2 \left( \begin{pmatrix} 50 \\ 50 \end{pmatrix}, \begin{pmatrix} 625 & 312.5 \\ 312.5 & 625 \end{pmatrix} \right)$$

- Test designed to have a mean of 50
- True mean constrained to be between 0 and 100; $\mu \pm 2\sigma = (0, 100)$ implies $\sigma^2 = (50/2)^2 = 625$.
- Prior correlation is 0.50
Priors Continued

- Prior for the precision matrix

\[ \phi \sim \text{Wishart}_2\left(4, \begin{pmatrix} 625 & 312.5 \\ 312.5 & 625 \end{pmatrix}^{-1}\right) \]

- loosely centered at the same covariance as in the normal mean prior distribution.

- \( \nu_0 \) is a prior degrees of freedom

For the full conditional distributions, we just need to find the updated hyperparameters (see Hoff)
multmodel=function() {
  for( i in 1 : N ) {
    Y[i,1:2] ~ dmnorm(mu[], Phi[],)
  }
  mu[1:2] ~ dmnorm(mu0[], prec[],)
  Phi[1:2, 1:2] ~ dwish(Phi0[,], nu0 )
  Sigma[1:2,1:2] <- inverse(Phi[],)
  rho <- Sigma[1,2]/sqrt(Sigma[1,1]*Sigma[2,2])
  Ynew[1:2] ~ dmnorm(mu[], Phi[],)
}
Joint Distribution of $\mu$
Predictive Distribution of a new student $Y$
Rao-Blackwellization

Monte Carlo estimates are noisy! Can do better by using the idea of “Rao-Blackwellization”

- $E_{MC}[g(\theta_1) \mid Y] = \frac{1}{T} \sum_t g(\theta_1^{(t)})$
- Iterated Expectations:

$$E[g(\theta_1) \mid Y] = E_{\theta_2} E_{\theta_1 \mid \theta_2}[g(\theta_1) \mid Y, \theta_2] = E_{\theta_2}[\tilde{g}(\theta_2) \mid Y] \approx \frac{1}{T} \sum_t \tilde{g}(\theta_2^{(t)}) \equiv E_{RB}[g(\theta_1) \mid Y]$$

- Calculate inner conditional expectation analytically
- May reduce variance over MCMC average (Liu et al 1996)
- Motivated by the classical Rao Blackwell Theorem that says that taking any estimator and conditioning on a sufficient statistic will reduce its mean squared error. (proof in Stat 215)
Density estimation

Rather than use a histogram estimate of the density, use a Raobackwell estimate. We know the posterior density of $\mu$ given $\Phi$ and $Y$, $p(\mu \mid \Sigma, Y)$

$$p(\mu \mid Y, \Phi) = \frac{1}{2\pi} \frac{p/2}{|P_n|^{1/2}} \exp \left\{ -\frac{1}{2}(\mu - \mu_n)'P_n(\mu - \mu_n) \right\}$$

where

- prior precision $P_0 = S_0^{-1}$
- posterior precision $P_n = P_0 + n\Phi$
- posterior mean $\mu_n = P_n^{-1}(P_0\mu_0 + \Phi n\bar{Y})$

Use MC average over draws of $\Phi^{(1)}, \ldots, \Phi^{(T)}$ to integrate out $\Phi$ to obtain the marginal

$$\hat{p}(\mu \mid Y) = \frac{1}{T} \sum_t p(\mu \mid Y, \Phi^{(t)})$$
Alternative Priors on $\Sigma$

No reason to restrict attention to semi-conjugate prior distributions, as long as the prior on $\Sigma$ lead to positive definite matrices. Another possibility:

- $1/\sigma_j^2 \mid \sigma^2 \overset{iid}{\sim} \text{gamma}(\nu/2, \nu/2\sigma_0^2)$
- $\sigma \sim C_+(0, 1)$ (half Cauchy)
- $\rho \sim B(2, 2)$

Requires Metropolis-Hastings algorithms as full conditionals distributions are not recognizable. Automatic in WinBugs.